IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | | | |
|---|---|---|---|
| In re Application of | : Joachim FRANK et al. | | |
| Serial No. | : Unassigned | Examiner : | Unassigned |
| Filed | : Herewith | Art Unit : | Unassigned |
| For | : METHOD AND APPARATUS FOR THE AUTOMATIC SEPARATING AND INDEXING OF MULTI-SPEAKER CONVERSATIONS | | |

November 30, 2001

## PRELIMINARY AMENDMENT

Assistant Commissioner for Patents
Washington, DC 20231

Sir:

Preliminary to the examination thereof, please amend the above-identified

application as follows:

**Delete all paragraphs of the abstract (Pages 20-21) and replace them with the**

**following paragraph:**

-- Disclosed are a method and apparatus for processing a continuous audio stream

containing human speech in order to locate a particular speech-based transaction in the

audio stream, applying both known speaker recognition and speech recognition

techniques. Only the utterances of a particular predetermined speaker are transcribed thus

providing an index and a summary of the underlying dialogue(s). In a first scenario, an

incoming audio stream, e.g. a speech call from outside, is scanned in order to detect audio

segments of the predetermined speaker. These audio segments are then indexed and only

the indexed segments are transcribed into spoken or written language. In a second

scenario, two or more speakers located in one room are using a multi-user speech

recognition system (SRS). For each user there exists a different speaker model and

optionally a different dictionary or vocabulary of words already known or trained by the

speech or voice recognition system.--

**Please cancel Claims 1-18, without prejudice, and add the following new Claims:**

--19. (New) A method of processing a continuous audio stream containing

human speech related to at least one particular transaction, comprising the steps of:

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

transcribing at least part of the continuous audio stream if a predetermined

speaker is recognized.--

--20. (New) A method according to claim 19, comprising a further step of

protocolling time information for detected speaker changes.

--21. (New) A method according to claim 19, wherein the step of detecting a speaker change and/or the step of performing a speaker recognition is/are preceded by a further step of detecting non-speech boundaries between continuous speech segments.--

--22. (New) A method according to claim 19, wherein the step of detecting a speaker change is accomplished by use of at least one characteristic audio feature, in particular features derived from the spectrum of the audio signal.--

--23. (New) A method according to claim 19, wherein the step of performing a speaker recognition involves the particular steps of calculating a speaker signature from the audio stream and comparing the calculated speaker signature with at least one known speaker signature.--

--24. (New) A method according to claim 19 for use in a speech recognition or voice control system comprising at least two speaker-specific speaker models and/or dictionaries, wherein interchanging between the at least two speaker-specific dictionaries is dependent on the detected speaker change and the corresponding recognized speaker.--

--25. (New) A method of processing a continuous audio stream containing human speech related to at least one particular transaction, comprising the steps of:

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

indexing the audio stream with respect to the detected speaker change if a predetermined speaker is recognized.--

--26. (New) A method according to claim 25, comprising a further step of protocolling time information for detected speaker changes.--

--27. (New) A method according to claim 25, wherein the step of detecting a speaker change and/or the step of performing a speaker recognition is/are preceded by a further step of detecting non-speech boundaries between continuous speech segments.--

--28. (New) A method according to claim 25, wherein the step of detecting a speaker change is accomplished by use of at least one characteristic audio feature, in particular features derived from the spectrum of the audio signal.--

--29. (New) A method according to claim 25, wherein the step of performing a speaker recognition involves the particular steps of calculating a speaker signature from the audio stream and comparing the calculated speaker signature with at least one known speaker signature.--

--30. (New) A method according to claim 25 for use in a speech recognition or voice control system comprising at least two speaker-specific speaker models and/or dictionaries, wherein interchanging between the at least two speaker-specific dictionaries is dependent on the detected speaker change and the corresponding recognized speaker.—

--31. (New)  An apparatus for processing a continuous audio stream containing human speech related to at least one particular transaction, comprising:

a predeterminer which predetermines at least one speaker;

a detector which detects speaker changes in the audio stream;

a recognizer which recognizes the predetermined speaker in the audio stream; and

an initiator which initiates transcription of at least part of the audio stream in case of a detected speaker change and a recognized predetermined speaker.--

--32. (New)  An apparatus according to claim 31, further comprising a detector which detects non-speech boundaries between continuous speech segments.--

--33. (New)  An apparatus according to claim 31, further comprising a scanner which automatically scans a continuous audio record, in particular a continuous audio stream recorded on a data or a signal carrier, and for detecting speaker changes in the continuous audio record.--

--34. (New)  An apparatus according to claim 31, further comprising a monitor which continuously monitors a real-time continuous audio stream and performing the steps of

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

transcribing at least part of the continuous audio stream if a predetermined

speaker is recognized.--

--35.  (New)  An apparatus according to claim 31, further comprising a monitor

which continuously monitors a real-time continuous audio stream and performing the

steps of

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

indexing the audio stream with respect to the detected speaker change if a

predetermined speaker is recognized.--

--36.  (New)  An apparatus according to claim 31, further comprising a logging

device which protocols time information for the at least one detected speaker change.--

--37.  (New)  An apparatus according to claim 31, comprising a marking device

which marks at least the beginning of a detected speech segment related to a

predetermined speaker.--

--38. (New) An apparatus according to claim 31, comprising data base which stores speech signatures for at least two speakers.--

--39. (New) An apparatus for processing a continuous audio stream containing human speech related to at least one particular transaction, comprising:

a predeterminer which predetermines at least one speaker;

a detector which detects speaker changes in the audio stream;

a recognizer which recognizes the predetermined speaker in the audio stream; and

an indexer for indexing the audio stream dependent on a detected speaker change and a recognized predetermined speaker.--

--40. (New) An apparatus according to claim 39, further comprising a detector which detects non-speech boundaries between continuous speech segments.--

--41. (New) An apparatus according to claim 39, further comprising a scanner which automatically scans a continuous audio record, in particular a continuous audio stream recorded on a data or a signal carrier, and for detecting speaker changes in the continuous audio record.--

--42. (New) An apparatus according to claim 39, further comprising a monitor which continuously monitors a real-time continuous audio stream and performing the steps of

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

transcribing at least part of the continuous audio stream if a predetermined

speaker is recognized.--

--43. (New) An apparatus according to claim 39, further comprising a monitor

which continuously monitors a real-time continuous audio stream and performing the

steps of

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

indexing the audio stream with respect to the detected speaker change if a

predetermined speaker is recognized.--

--44. (New) An apparatus according to claim 39, further comprising a logging

device which protocols time information for the at least one detected speaker change.--

--45. (New) An apparatus according to claim 39, comprising a marking device which marks at least the beginning of a detected speech segment related to a predetermined speaker.--

--46. (New) An apparatus according to claim 39, comprising data base which stores speech signatures for at least two speakers.--

--47. (New) A speech recognition or voice control system processing an incoming audio stream and having at least two speaker models and/or speaker-specific dictionaries, comprising:

a detector which detects a speaker change in the incoming audio stream;

a gatherer which gathers speaker-specific information and for comparing the gathered speaker-specific information with corresponding speaker-specific information of at least one predetermined speaker thus recognizing the at least one predetermined speaker; and

an interchanger which interchanges between the at least two speaker-specific dictionaries dependent on the detected speaker change and the corresponding recognized speaker.--

--48. (New) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for

processing a continuous audio stream containing human speech related to at least one

particular transaction, said method comprising the steps of:

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

transcribing at least part of the continuous audio stream if a predetermined

speaker is recognized.--

--49. (New)  A program storage device readable by machine, tangibly embodying

a program of instructions executable by the machine to perform method steps for

processing a continuous audio stream containing human speech related to at least one

particular transaction, said method comprising the steps of:

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

indexing the audio stream with respect to the detected speaker change if a

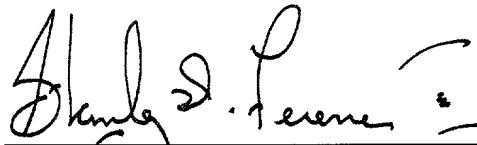predetermined speaker is recognized.--

## REMARKS

The abstract has been deleted and a new abstract has been substituted therefor. The new abstract complies with the requirements of 37 C.F.R. § 1.72.

Claims 1-18 as filed in Europe have been canceled. New Claims 19-49 have been added. Claims 19-49 generally correspond to Claims 1-18, however, the newly added claims appear in the format typically used in U.S. practice and do not include multiple dependencies. No change in scope is intended.

A marked-up version of the changes made by this Preliminary Amendment is attached.

Respectfully submitted,

Stanley D. Ference III
Registration No. 33,879

FERENCE & ASSOCIATES
129 Oakhurst Road
Pittsburgh, Pennsylvania 15215
(412) 781-7386
(412) 781-8390 - Facsimile

Attorneys for Applicants

## VERSION WITH MARKINGS TO SHOW CHANGES MADE

**The abstract has been deleted and the following substituted therefor:**

--Disclosed are a method and apparatus for processing a continuous audio stream containing human speech in order to locate a particular speech-based transaction in the audio stream, applying both known speaker recognition and speech recognition techniques. Only the utterances of a particular predetermined speaker are transcribed thus providing an index and a summary of the underlying dialogue(s). In a first scenario, an incoming audio stream, e.g. a speech call from outside, is scanned in order to detect audio segments of the predetermined speaker. These audio segments are then indexed and only the indexed segments are transcribed into spoken or written language. In a second scenario, two or more speakers located in one room are using a multi-user speech recognition system (SRS). For each user there exists a different speaker model and optionally a different dictionary or vocabulary of words already known or trained by the speech or voice recognition system.--

**Claims 1-18 have been cancelled and the following Claims 19-49 have been added:**

--19. (New) A method of processing a continuous audio stream containing human speech related to at least one particular transaction, comprising the steps of:

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

transcribing at least part of the continuous audio stream if a predetermined

speaker is recognized.--

--20. (New) A method according to claim 19, comprising a further step of

protocolling time information for detected speaker changes.

--21. (New) A method according to claim 19, wherein the step of detecting a

speaker change and/or the step of performing a speaker recognition is/are preceded by a

further step of detecting non-speech boundaries between continuous speech segments.--

--22. (New) A method according to claim 19, wherein the step of detecting a

speaker change is accomplished by use of at least one characteristic audio feature, in

particular features derived from the spectrum of the audio signal.--

--23. (New) A method according to claim 19, wherein the step of performing a

speaker recognition involves the particular steps of calculating a speaker signature from

the audio stream and comparing the calculated speaker signature with at least one known

speaker signature.--

--24. (New) A method according to claim 19 for use in a speech recognition or

voice control system comprising at least two speaker-specific speaker models and/or

dictionaries, wherein interchanging between the at least two speaker-specific dictionaries

is dependent on the detected speaker change and the corresponding recognized speaker.--

--25. (New)  A method of processing a continuous audio stream containing

human speech related to at least one particular transaction, comprising the steps of:

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

indexing the audio stream with respect to the detected speaker change if a

predetermined speaker is recognized.--

--26. (New)  A method according to claim 25, comprising a further step of

protocolling time information for detected speaker changes.--

--27. (New)  A method according to claim 25, wherein the step of detecting a

speaker change and/or the step of performing a speaker recognition is/are preceded by a

further step of detecting non-speech boundaries between continuous speech segments.--

--28. (New)  A method according to claim 25, wherein the step of detecting a

speaker change is accomplished by use of at least one characteristic audio feature, in

particular features derived from the spectrum of the audio signal.--

--29. (New) A method according to claim 25, wherein the step of performing a speaker recognition involves the particular steps of calculating a speaker signature from the audio stream and comparing the calculated speaker signature with at least one known speaker signature.--

--30. (New) A method according to claim 25 for use in a speech recognition or voice control system comprising at least two speaker-specific speaker models and/or dictionaries, wherein interchanging between the at least two speaker-specific dictionaries is dependent on the detected speaker change and the corresponding recognized speaker.—

--31. (New) An apparatus for processing a continuous audio stream containing human speech related to at least one particular transaction, comprising:

a predeterminer which predetermines at least one speaker;

a detector which detects speaker changes in the audio stream;

a recognizer which recognizes the predetermined speaker in the audio stream; and

an initiator which initiates transcription of at least part of the audio stream in case of a detected speaker change and a recognized predetermined speaker.--

--32. (New) An apparatus according to claim 31, further comprising a detector which detects non-speech boundaries between continuous speech segments.--

--33. (New) An apparatus according to claim 31, further comprising a scanner which automatically scans a continuous audio record, in particular a continuous audio stream recorded on a data or a signal carrier, and for detecting speaker changes in the continuous audio record.--

--34. (New) An apparatus according to claim 31, further comprising a monitor which continuously monitors a real-time continuous audio stream and performing the steps of

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

transcribing at least part of the continuous audio stream if a predetermined speaker is recognized.--

--35. (New) An apparatus according to claim 31, further comprising a monitor which continuously monitors a real-time continuous audio stream and performing the steps of

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

indexing the audio stream with respect to the detected speaker change if a predetermined speaker is recognized.--

--36. (New) An apparatus according to claim 31, further comprising a logging device which protocols time information for the at least one detected speaker change.--

--37. (New) An apparatus according to claim 31, comprising a marking device which marks at least the beginning of a detected speech segment related to a predetermined speaker.--

--38. (New) An apparatus according to claim 31, comprising data base which stores speech signatures for at least two speakers.--

--39. (New) An apparatus for processing a continuous audio stream containing human speech related to at least one particular transaction, comprising:

a predeterminer which predetermines at least one speaker;

a detector which detects speaker changes in the audio stream;

a recognizer which recognizes the predetermined speaker in the audio stream; and

an indexer for indexing the audio stream dependent on a detected speaker change and a recognized predetermined speaker.--

--40. (New) An apparatus according to claim 39, further comprising a detector which detects non-speech boundaries between continuous speech segments.--

--41. (New) An apparatus according to claim 39, further comprising a scanner which automatically scans a continuous audio record, in particular a continuous audio stream recorded on a data or a signal carrier, and for detecting speaker changes in the continuous audio record.--

--42. (New) An apparatus according to claim 39, further comprising a monitor which continuously monitors a real-time continuous audio stream and performing the steps of

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

transcribing at least part of the continuous audio stream if a predetermined speaker is recognized.--

--43. (New) An apparatus according to claim 39, further comprising a monitor which continuously monitors a real-time continuous audio stream and performing the steps of

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

indexing the audio stream with respect to the detected speaker change if a predetermined speaker is recognized.--

--44. (New) An apparatus according to claim 39, further comprising a logging device which protocols time information for the at least one detected speaker change.--

--45. (New) An apparatus according to claim 39, comprising a marking device which marks at least the beginning of a detected speech segment related to a predetermined speaker.--

--46. (New) An apparatus according to claim 39, comprising data base which stores speech signatures for at least two speakers.--

--47. (New) A speech recognition or voice control system processing an incoming audio stream and having at least two speaker models and/or speaker-specific dictionaries, comprising:

a detector which detects a speaker change in the incoming audio stream;

a gatherer which gathers speaker-specific information and for comparing the

gathered speaker-specific information with corresponding speaker-specific information of

at least one predetermined speaker thus recognizing the at least one predetermined

speaker; and

an interchanger which interchanges between the at least two speaker-specific

dictionaries dependent on the detected speaker change and the corresponding recognized

speaker.--

--48.  (New)  A program storage device readable by machine, tangibly embodying

a program of instructions executable by the machine to perform method steps for

processing a continuous audio stream containing human speech related to at least one

particular transaction, said method comprising the steps of:

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

transcribing at least part of the continuous audio stream if a predetermined

speaker is recognized.--

--49.  (New)  A program storage device readable by machine, tangibly embodying

a program of instructions executable by the machine to perform method steps for

processing a continuous audio stream containing human speech related to at least one

particular transaction, said method comprising the steps of:

digitizing the continuous audio stream;

detecting a speaker change in the digitized audio stream;

performing a speaker recognition if a speaker change is detected; and

indexing the audio stream with respect to the detected speaker change if a predetermined

speaker is recognized.--